



Digital Collections: Making it Happen

Hema Ramachandran
Ed Sponsler
Jim O'Donnell,
Caltech Library System
SCELC, September 13 2002, Caltech

Caltech Digital Collections

- <http://library.caltech.edu/digital/>
 - What do we have?
 - Where did these come from?
 - How did we get here?
 - Why digital collections?



Identifying projects

- No barriers to implementation
 - Established collections
 - Documents already in some digital format
 - Blanket publishing license available
 - Describable with minimal metadata: author/title/abstract



Prep work

- Develop archive policy documents
 - Harvesting policy
 - Submission policy
 - Instructions for new submissions
 - Archive subject description (the About page/scope)



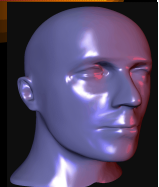
In-house cooperation

- *Extremely* important
 - Assistance of other departments
 - Library Information Technology Group
 - Document Delivery
 - Keeping track of documents
 - Develop procedures
 - Quality Control



caltechCSTR

- Established concept
- Dept. Chair is strong advocate
- Department needing assistance
- Documents already created
 - (electronic and print)
- Single blanket permission to
 - publish and for retrospective conversion
- Rejoining NCSTRL



caltechCSTR (continued)

- Migrating existing electronic files
- Conversion from print to digital (scanning old reports)
- Submission of new reports



Submission of New Reports

- Issuing report numbers
- Designing author submission interface
- Outlining steps for author submission
- Approval process
- Copyright
- Document preparation
- Promoting
- CaltechCSTR



caltechEERL

- Earthquake Engineering Research Laboratory
- Collection of published paper technical reports
- Blanket permission for conversion
- Collection more than just EERL Reports
- In most cases loose-leaf masters available



caltechEERL -- processing

- Digitize backfiles into TIFF files
- Scan TIFF to obtain abstract
- OCR abstract
- Complete metadata
- Submit the files
- Deposit zipped TIFF file into Document Source Archive
- Store file cabinets with paper masters



CAV2001

- Browse by Session helpful to attendees
- International Submissions: 90% accuracy
- 10% needed work
 - Trouble creating PDF from TeX
 - PDF Font Issues
 - Grad student liaison
- Exclusively Web-based Digital Publication
- Cataloged in OCLC



caltechLIB

- Bibliography of CLS staff professional publications
- Good first archive
 - No need to obtain permission to publish
 - No need to solicit 'buy-in'
 - More control over document format / preparation
 - Opportunity to fiddle around with the EPrints software



Open Archives Initiative

- What is the OAI-PMH?
 - Protocol for Metadata Harvesting
 - A Service Provider harvests from Data Providers
- EPrints is an OAI-PMH Data Provider
- Built around a basic subset of Dublin Core
 - Author / Title / Abstract / Date / URL
 - Thus able to support a wide variety of archives
- <http://www.openarchives.org>

Impact of OAI

- Register your OAI compliant archive
- Immediate benefits
 - DP9: For Web Crawlers
 - ARC: Federated Searching
 - Additional services as they come
- Future developments at Caltech
 - Federated Searching across Caltech archives



Persistent URLs (PURLs)

- A permanent URL to a record
- Beneficial in:
 - Document citation
 - Document header/title page
- Developed In House – Eprints compatible



Just do it!

- Making a start is important
- Many issues to consider, but not necessarily all at once
- Be aware of author's needs
- Archives attract more archives!
- Keep expectations realistic



Don't get bogged down

- Metadata
- Subject access
- Document enhancements



The Caltech experience

- Without grant funding
- Using in-house professional staff
- Simultaneous development and production
- Entrepreneurial
- Team work
- Integrating with existing work

